Sobre la existencia y el cálculo de ceros de funciones regulares

POR J. M. SORIANO ARBIZU

Recibido: 18 de noviembre de 1988

Presentado por el académico correspondiente D. Antonio de Castro Brzezicki

Abstract

For a map f of class two a sufficient condition is obtained to secure the existence of a zero of f in the simplex s if a piecewise linear aproximation θ of f has a cero in s.

Application: for a smooth map F: $D \subseteq \mathbf{C}^n \longrightarrow \mathbf{C}^n$ a noniterative algorithm is constructed of obtain the zeros of F using the usual homotopy H in $D \times [0, 1]$ and a piecewise linear approximation θ relative to a triangulation K of $D \times [0, 1]$. An implemention of the procedure is cited.

Resumen

Para una aplicación f de clase dos se obtiene una condición suficiente para asegurar la existencia de un cero de f en el símplex s si una aproximación lineal a trozos θ de f tiene un cero en s.

Aplicación: para una aplicación regular F: $D \subseteq \mathbf{C}^n \longrightarrow \mathbf{C}^n$ se construye un algolritmo no iterativo para calcular sus ceros, utilizando la homotopía usual H en $D \times [0, 1]$ y una aproximación lineal a trozos θ relativa a una triangulación K de $D \times [0, 1]$. Se cita una aplicación del procedimiento.

INTRODUCCION

El método seguido consiste en sustituir en un cierto dominio una función f por otra lineal y próxima a ella θ y utilizar las propiedades del grado topológico [5] que nos permiten obtener una condición suficiente de existencia de ceros de la función problema en un cierto símplex si lo tiene la función aproximante θ. Esta última se caracteriza por tener igual valor numérico que f en los vértices de un sìmplex de diámetro suficientemente pequeño, y se extiende baricéntricamente a los restantes puntos de éste.

En el apartado 2 se construye un algoritmo no iterativo para calcular los ceros de cierta clase de funciones. En primer lugar construimos una homotopía H con la función problema y otra auxiliar cuyos ceros se conocen. Los teoremas 2-1, 2-2, 2-3 tienen por finalidad conocer la naturaleza de $H^{-1}(0)$. Vemos que son caminos que terminan en un cero de la función problema y empiezan en un cero de la función auxiliar.

Las secciones 2-4 y 2-5 están dirigidas a aproximar H.

En [3], [8] figuran varias construcciones de homotopías y se utilizan métodos iterativos para calcular cada cero, pero sin demostrar la efectiva existencia del mismo. Nosotros demostramos esta existencia y además la técnica utilizada permite un cálculo más rápido.

*1. **Teorema.***– Sea f: $D \subseteq \mathbf{R}^n \to \mathbf{R}^n$ una función de clase dos y θ: $D \subseteq \mathbf{R}^n \to \mathbf{R}^n$ una aproximación lineal a trozos de f relativa a una triangulación de $\mathbf{R}^n$. Entonces, si θ tiene un cero en el n-símplex s, f también tiene un cero en s, si se verifica

$$2 \times M \times \xi^2 < d_\infty (0, \theta (\partial s));$$

siendo M una cota superior para la norma de la segunda derivada de f en s, y ξ el diámetro de s.

Demostración.– Si θ tiene un cero en s y cero no es un valor crítico, se deduce de la definición de grado topológico que $d(\theta, s, 0) = 1$.

Aplicamos las dos siguientes propiedades de grado topológico: 1) si $d(f, s, 0) \neq 0 \Rightarrow f^{-1}(0) \neq \emptyset$; 2) $d(., s, 0)$ es constante sobre $\{f\} \in C(\bar{s})$ si

$$\sup_{X \in s} |f(X) - \theta(X)|_\infty < r,$$

con $r = d_\infty (0, \theta(\partial s))$ [5].

Calculemos una cota para $\max\limits_{X \in \bar{s}} |f(X) - \theta(X)|_\infty$,

$$\max_{X \in \bar{s}} |f(X) - \theta(X)|_\infty =$$

$$= \max_{X \in \bar{s}} (|f_1(X) - \theta_1(X)| , ..., |f_n(X) - \theta_n(X)|)$$

Para esto:

a) Suponemos que

$$\max_{X \in \bar{s}} |f(X) - \theta(X)|_\infty = |f_j(X^*) - \theta_j(X^*)| ;$$

$$X^* \in \bar{s}, \quad j \in 1, ..., n.$$

b) Construimos la función f^*

$$f^*\colon D \subseteq \mathbf{R}^n \to \mathbf{R}^n;$$

$$f_k^*(X) = f_k^*(X, Y^*) = f_k(X) + \sum_{i=1}^{n} ((\partial f_k(Y^*) / \partial X_i) \times h_i,$$

$k = 1, ..., n$ con $X \in \bar{s}$, $Y^* \in s$ conocido; $s = \langle Y^0, ..., Y^n \rangle$;

$$X = \sum_{i=0}^{n} \lambda_i \times Y^i; \qquad \sum_{i=0}^{n} \lambda_i = 1; \qquad \lambda_i \geqslant 0, \qquad i = 0, ..., n;$$

$$h_k = X_k - Y_k^* = \sum_{i=0}^{n} \lambda_i \times Y_k^i - Y_k^*, \qquad k = 1, ..., n.$$

c) Maximizamos $|f_j^* (X) - \theta_j (X)|$ para $X \in \bar{s}$;

$$\max_{X \in \bar{s}} |f_j^* (X) - \theta_j (X)| =$$

$$= \max_{X \in \bar{s}} \left| \sum_{k=1}^{n} \lambda_k \times \left\{ \left[\sum_{i=1}^{n} ((\partial f_j (Y^*) / \partial X_i) \times Y_i^k \right] - Y_j^k \right\} + \right.$$

$$\left. + \left\{ f_j (Y^*) - \left[\sum_{i=1}^{n} ((\partial f_j (Y^*) / \partial X_i)) \times Y_i^* \right] \right\} \right| ;$$

que escrito en forma reducida es

$$\max_{X \in \bar{s}} |f_j^* (X) - \theta_j (X)| = \max_{X \in \bar{s}} |A_0 \times \lambda_0 + ... + A_n \times \lambda_n + B|$$

con

$$A_k = \left[\sum_{i=1}^{n} ((\partial f_j (Y^*) / \partial X_i) \times Y_i^k \right] - Y_j^k; \qquad k = 0, ..., n;$$

$$B = f_j (Y^*) - \left[\sum_{i=1}^{n} ((\partial f_j (Y^*) / \partial X_i) \times Y_i^* \right].$$

Calcular este máximo equivale a calcular el máximo de $|A_0 \times \lambda_0 + ... + A_n \times \lambda_n + B|$ con las condiciones

$$\sum_{i=0}^{n} \lambda_i = 1; \qquad \lambda_i \geqslant 0, \qquad i = 0, ..., n.$$

Así pues, calculamos

$$\text{máximo de} \begin{cases} |\text{máximo de } \{A_0, ..., A_n\} + B| \\ |\text{mínimo de } \{A_0, ..., A_n\} + B|; \end{cases}$$

por tanto, este máximo ocurre para un $\lambda_i = 1$, $i = 0, ..., n$.

Esto implica que

$$\max_{X \in \bar{s}} |f_j^* (X) - \theta_j (X)| = |f_j^* (Y^i) - f_j (Y^i)|,$$

porque

$$\theta_j (Y^i) = f_j (Y^i).$$

Por ser $|f^* (X) - f(X)|_\infty \leqslant M \times \xi^2$, tenemos

$$\max_{X \in \bar{s}} |f_j^* (X) - \theta_j (X)| \leqslant M \times \xi^2.$$

d) Calculamos por último una cota para $\max_{X \in \bar{s}} |f(X) - \theta (X)|_\infty$,

$$\max_{X \in \bar{s}} |f(X) - \theta (X)|_\infty =$$

$$= |f_j (X^*) - \theta_j (X^*)| \leqslant \max_{X \in \bar{s}} |f_j (X) - f_j^* (X)| +$$

$$+ \max_{X \in \bar{s}} |f_j^* (X) - \theta_j (X)| \leqslant 2 \times M \times \xi^2.$$

$$\sup_{X \in s} |f(X) - \theta (X)|_\infty \leqslant \max_{X \in \bar{s}} |f(X) - \theta (X)|_\infty.$$

Así, cuando $2 \times M \times \xi^2 < d_\infty (0, \theta (\partial s))$, aplicando las propiedades 2) y 1) sucesivamente, si hay un cero de θ en s, también hay un cero de f en s.

Observaciones.–

A) Para comprobar las hipótesis del teorema necesitamos conocer el valor de $d_\infty (0, \theta (\partial s))$ y un valor de M; el cálculo es sencillo; así, para calcular $d_\infty (0, \theta (\partial s))$ vemos:

1) Cada n puntos del conjunto $\{f(Y^0), ..., f(Y^n)\}$ ($Y^0, ..., Y^n$ pertenecen a un n-símplex de $\mathbf{R}^n$) determinan un hiperplano de $\mathbf{R}^n$ que contiene el cierre de un $(n - 1)$-símplex perteneciente a $\theta (\partial s)$; así obtenemos $(n + 1)$-hiperplanos que contienen $\overline{\theta (\partial s)}$. La ecuación de uno de estos hiperplanos tiene la forma:

$$\begin{vmatrix} 1 & 1 & \dots & 1 \\ X & f(Y^{i_0}) & \dots & f(Y^{i_m}) \end{vmatrix} = 0,$$

ó

$$A_1^i \times X_1 + \dots + A_m^i \times X_m + A_{m+1}^i = 0 \equiv \pi_i$$

2) $d_\infty (0, \pi_i) = \text{máx} (|A_1^i| \times |A_{n+1}^i| : |A^i|_2^2, ..., |A_n^i| \times |A_{n+1}^i| : |A^i|_2^2,$

con $A^i = (A_1^i, ..., A_n^i)$.

3) Así

$$d_\infty(0, \theta(\partial s)) = \text{mín}\,\{d_\infty(0, \pi_1), ..., d_\infty(0, \pi_{n+1})\}.$$

Para calcular el valor de M:

$$|f(X) - f(Y^i) - f'(Y^i) \times (X - Y^i)|_\infty \leqslant$$

$$\leqslant (1/2) \times |f''(Y^i + t \times (X - Y^i)|_\infty \times |X - Y^i|_\infty^2 \leqslant M \times \xi^2,$$

con $X \in s$, $t \in [0, 1]$, $\sup |f''(Y^i + t \times (X - Y^i)|_\infty = M$.

$$\sup |f''(Y^i + t(X - Y^i)|_\infty \leqslant$$

$$\leqslant \sup |k^t \times \text{Hessiano}\,[f_1(Y^i + t \times (X - Y^i)] \times h, ...,$$

$$k^t \times \text{Hessiano}\,[f_n(Y^i + t \times (X - Y^i)] \times h|_\infty,$$

con $0 \leqslant t \leqslant 1$, $|h|_\infty = |k|_\infty = 1$.

Con la triangulación K_1 [8] y eligiendo $Y^i = Y^0$; $h_j, k_j \in [0, 1]$, $j = 1, ..., n+1$; y diámetro de malla ξ, obtenemos que cada componente de este vector es menor o igual que:

$$\text{máx}\left\{\begin{array}{l} \sum\limits_{j,k} (\partial^2 f_i(Y^0 + p)) / \partial X_j \partial X_k; \\ (\partial^2 f_i(Y^0 + p)) / \partial X_j \partial X_k > 0; \quad p_i \in [0, \xi]. \\ \\ \left| \sum\limits_{m,n} (\partial^2 f_i(Y^0 + p) / \partial X_m \partial X_n \right|; \\ (\partial^2 f_i(Y^0 + p)) / \partial X_m \partial X_n < 0. \end{array}\right.$$

B) El teorema 1 completa algunos métodos para localizar los ceros de una función vectorial f. Por ejemplo, completa el método dado por Hsu y Zhu en [4] que encuentra ceros de una aproximación lineal a trozos θ de f relativa a una triangulación K en un cierto dominio de interés. Esto es, encuentra ceros aproximados de f ó puntos que están próximos a sus imágenes por f, pero no necesariamente próximos a un cero.

2. Algoritmo.– Para resolver $F^*(z) = 0$, F^*: $D \subseteq \mathbf{C}^n \longrightarrow \mathbf{C}^n$; siendo $F^* \in C^\infty(D)$ con un número finito de ceros, identificamos $\mathbf{C}^n$ con $\mathbf{R}^{2\times n}$ y empezamos con el sistema soluble $G^*(z) = 0$, G^*: $D \subseteq \mathbf{C}^n \longrightarrow \mathbf{C}^n$; G^* es $C^\infty(D)$.

Construimos la homotopía

$$H^*: D \times [0, 1] \longrightarrow \mathbf{C}^n;$$

$$H^* = (H_1^*, ..., H_n^*), \qquad H^*(z, t) = (1 - t) \times G^*(z) + t \times F^*(z),$$

y resolvemos

$$H^*(z, t) = 0; \qquad H^*(z, 0) = G^*(z), \qquad H^*(z, 1) = F^*(z).$$

Sean

$$H: D \times [0, 1] \longrightarrow \mathbf{R}^{2 \times n},$$

con $H = (\text{Re } H_1^*, \text{ Imag } H_1^*, ..., \text{ Imag } H_n^*)$; variable de $H(u, t)$, $u = (\text{Re } z_1, \text{ Imag } z_1, ..., \text{ Imag } z_n)$.

$$F: D \subseteq \mathbf{R}^{2 \times n} \longrightarrow \mathbf{R}^{2 \times n},$$

con $F = (\text{Re } F_1^*, \text{ Imag } F_1^*, ..., \text{ Imag } F_n^*)$.

$$G: D \subseteq \mathbf{R}^{2 \times n} \longrightarrow \mathbf{R}^{2 \times n},$$

con $G = (\text{Re } G_1^*, \text{ Imag } G_1^*, ..., \text{ Imag } G_n^*)$.

Así, resolver $H^*(z, t) = 0$, y $H(u, t) = 0$, son problemas equivalentes.

Suponemos que cero es un valor regular para H. Esto no es substancialmente restrictivo en vista del corolario de Brown [6] del teorema de Sard.

Estudiemos la naturaleza de $H^{-1}(0)$ con la ayuda de los siguientes teorema [6]; con $m = 2 \times n + 1$, $p = 2 \times n$, $M = D \times [0, 1]$, y $N = R^{2 \times n}$.

2.1. Teorema.– Si H es una aplicación regular $H: M \longrightarrow N$ desde una m-variedad con borde M a una p-variedad N, donde $m > p$, e Y es un valor regular, para H y para la restricción $H | \partial M$, sigue que $H^{-1}(Y)$ es una $(m - p)$-variedad regular con borde. Además, el borde $\partial(\mathrm{H}^{-1}(Y))$ es la intersección de $H^{-1}(Y)$ con ∂M, y su dimensión es $m - p - 1$.

2.2. Teorema.– Cualquier 1-variedad regular conexa es difeomorfa a un círculo S^1 ó a algún intervalo real; esto es, es un lazo o un camino.

Para calcular los ceros de F^* seguimos las componentes conexas de $H^{-1}(0)$ desde $t = 0$ hasta $t = 1$. Esto es posible porque $H^{-1}(0)$ es una unión de caminos y lazos, según se deduce de los dos teoremas anteriores.

Sea s el parámetro longitud de las componentes conexas de $H^{-1}(0)$, $\{u, t\}$; $H(u(s), t(s)) = 0$. Diferenciando,

$$\sum_{i=1}^{2 \times n} (\partial H / \partial u_i) \times \dot{u}_i + (\partial H / \partial t) \times \dot{t} = 0,$$

sigue [2]

$$\dot{u}_i = (-1)^i \times \det (H'_{-i} (u, t)), \qquad i = 1, ..., n \tag{2}$$

$$\dot{t} = (-1)^{2 \times n + 1} \times \det (H'_{-2 \times n - 1} (u, t)). \tag{1}$$

2.3. Teorema.– Si H^* es holomorfa, entonces $\det (H'_{-2n-1} (u, t)) \geqslant 0$.
De (1) y el teorema 2-3 sigue que los caminos de $H^{-1}(0)$ son monónotonos en t y no hay lazos, si H es holomorfa.

Si elegimos G tal que ninguno de los caminos de $H^{-1}(0)$ diverja a infinito para $0 \leqslant t < 1$; cada camino que empieza en $(X^*, 0)$ con $G(X^*) = =0$ termina en $(X^{**}, 1)$, $F(X^{**}) = 0$. Así, resolviendo el problema de valores iniciales (1), (2) para cada una de las soluciones de $G(u) = 0$, obtenemos las soluciones de $F(u) = 0$ y quizás algún camino que diverge a infinito. [2] da una condición suficiente para que esto no ocurra.

2.4. Aproximación lineal a trozos.– Necesitamos unos dominios de $\mathbf{R}^{2 \times n} \times [0, 1]$ cuyos cierres contengan las componentes conexas de $H^{-1}(0)$, siendo cero un valor regular para H; tal que la intersección del cierre de estos dominios con el hiperplano $X_{2 \times n + 1} = 1$ sean entornos de las soluciones de nuestro problema. Para obtenerlos, construimos una triangulación regular K de $R^{2 \times n} \times [0, 1]$, tal que X^* que verifica $G(X^*) = 0$, esté contenido en el interior de uno de sus $2 \times n$-símplices y que el diámetro de sus $2 \times n$-símplices sea menor que el error admisible.

Construimos una aproximación lineal a trozos a H relativa a la triangulación K que llamamos θ.

$$\theta: D \times [0, 1] \longrightarrow \mathbf{R}^{2 \times n};$$

$$\theta(X) = \begin{cases} H(u, t) \text{ para } (u, t) \in \{\text{vértices de } K\} \\ \sum_{i=0}^{k} \lambda_i \times H(Y^i) \text{ para } X = (u, t) \in \langle Y^0, ..., Y^k \rangle, \end{cases}$$

siendo λ_i las coordenadas baricéntricas de X respecto de $\{Y^0, ..., Y^k\}$.

El estudio de $\theta^{-1}(0)$ es el usual en este tipo de algoritmos.

2.5. Construcción del algoritmo.– Seguimos una componente conexa de $H^{-1}(0)$, recorriendo una componente conexa de $\theta^{-1}(0)$. La malla de la triangulación se construye suficientemente fina para garantizar la aproximación deseable entre los ceros de ambas funciones; adoptando un diámetro de malla menor que el error deseable para las soluciones. En el primer paso conocemos la existencia de un cero de G en el interior de un $2 \times n$-símplex específico s, en el que también hay un cero de θ si se verifica que el rango de la matriz

$$\begin{pmatrix} 1 & \dots & 1 \\ H(Y^0) & \dots & H(Y^{2n}) \end{pmatrix}$$

es máximo; y $\lambda_i > 0$, $i = 0, \dots, 2 \times n$ para

$$\begin{pmatrix} 1 & \dots & 1 \\ H(Y^0) & \dots & H(Y^{2n}) \end{pmatrix} \times (\lambda_0, \dots, \lambda_{2n})^t = (1, 0, \dots, 0)^t;$$

$$s = \langle Y^0, \dots, Y^{2n} \rangle.$$

Suponemos ahora conocida la existencia de un cero de θ en un $2 \times n$-símplex. Tenemos dos posibilidades para construir un algoritmo: 1) seguir una componente conexa de $\theta^{-1}(0)$ y comprobar si el $2 \times n$-símplex s_F, que contiene la intersección de esta componente conexa de $\theta^{-1}(0)$ con el hiperplano $t = 1$, contiene un cero de F. 2) En cada $2n$-símplex que es intersecado por una componente conexa de $\theta^{-1}(0)$, s_i, comprobamos la existencia de un cero de H. En el caso negativo, volvemos al anterior $2 \times n$-símplex s_{i-1}, y cambiamos la malla de la triangulación.

Para realizar esta comprobación aplicamos el teorema 1.

En nuestro algoritmo hacemos esta comprobación en un $2 \times n$-símplex s de $R^{2 \times n + 1}$. Esto implica que s está contenido en un hiperplano; de donde sigue que H y θ aplican $R^{2 \times n}$ en $R^{2 \times n}$ como se requiere en el teorema.

Hemos programado en Fortran 77 y resuelto en un VAX 11/785 con este algoritmo sistemas del tipo

$$a_1X^2 + a_2XY + a_3 = 0; \qquad b_1Y^3 + b_2X + b_3 = 0;$$

$$a_i, b_i \in C, \qquad i = 1, 2, 3; \qquad X, Y \in \mathbf{C}$$

con el sistema auxiliar: $\{X^2 - 1 = 0;\ Y^3 - 1 = 0\}$.

Y del tipo:

$$a_1X^3 + a_2U + a_3 = 0; \qquad b_1Y^3 + b_2Z + b_3 = 0;$$

$$c_1Z^2 + c_2Y + c_3 = 0; \qquad d_1X + d_2Z + d_3U + d_4 = 0;$$

$$a_i, b_i, c_i, d_i \in \mathbf{C}, \qquad i = 1, \dots, 4; \qquad X, Y, Z, U \in \mathbf{C}$$

con el sistema auxiliar: $\{X^3 - 1 = 0,\ Y^3 - 1 = 0,\ Z^2 - 1 = 0,\ U - 1 = 0\}$.

DEFINICIONES

j-símplex abierto $s = \langle Y^0, ..., Y^j \rangle$ de $\mathbf{R}^n$ es el interior de la envoltura convexa de $(j+1)$-puntos afínmente independientes, llamados vértices [8].

K triangula a M, con M subconjunto convexo de $\mathbf{R}^n$, significa: a) K es una colección de n-símplices; b) los subsímplices de todos los n-símplices de K forman una partición de M [8].

Aproximación lineal a trozos θ a f; $f\colon D \subseteq \mathbf{R}^n \longrightarrow \mathbf{R}^n$ relativa a la triangulación K de D es la función θ definida por

$$\theta(X) = \begin{cases} f(x) & \text{si } X \text{ pertenece a los vértices de } K \\ \sum_{i=0}^{j} \lambda_i Y^i & \text{si } X \text{ pertenece al interior de } \langle Y^0, ..., Y^j \rangle; \end{cases}$$

con λ_i, $i = 0, ..., j$ las coordenadas baricéntricas de X respecto de $\{Y^0, ..., Y^j\}$, [8].

Sea s un abierto de $\mathbf{R}^n$; f una función de $\bar{s}$ en $\mathbf{R}^n$ que tiene una extensión de clase uno a algún abierto que contenga a $\bar{s}$. Sea cero un valor regular de f y que no pertenece a la imagen de la frontera de s por f. Llamamos grado de f respecto de cero en s, $d(f, s, 0)$ a $d(f, s, 0) = \Sigma J_f(X)$, con X perteneciente a $f^{-1}(0)$, [5].

BIBLIOGRAFIA

[1] BAZARAA, M. S. Y JARVIS, J. (1981): *Programación lineal y flujo en redes.* (Méjico: Limusa).

[2] GARCIA, C. B. Y ZANGWILL, W. I. (1981): *Pathways to solutions, fixed points and equilibria.* (Englewood Cliffs: Prentice-Hall).

[3] GARCIA, C. B. Y ZANGWILL, W. I. (1979): *Finding all solutions to polynomial system and other equations.* Math. Programming Vol. 16, pp. 159-176.

[4] HSU, C. S. Y ZHU, W. H. (1984): *A simplicial mapping method for locating the zeros of a function.* Quart. Appl. Math. Vol. 42-1, pp. 41-60.

[5] DEIMLING, K. (1984): *Non linear Functional Analysis.* (Berlin, Heidelberg, New York, Tokio: Springer Verlag).

[6] MILNOR, J. (1965): *Topology from the differentiable viewpoint.* (Charlottesville: University Press of Virginia).

[7] SCARF, H. (1967): *The approximation of fixed points of a continuous mapping.* SIAM J. Appl. Math. Vol. 15, pp. 1328-1343.

[8] TODD, J. (1976): *The computations of fixed points and applications.* Lecture Notes in Econom. and Math. Systems nº 124 (Berlin, Heidelberg, New York: SpringerVerlag).

Dpto. de Análisis Matemático
Facultad de Matemáticas
41012 Sevilla